The Astrophysical Journal, 724:526–538, 2010 November 20 doi:10.1088/0004-637X/724/1/526

BRIGHT SOURCE SUBTRACTION REQUIREMENTS FOR REDSHIFTED 21 cm MEASUREMENTS

A. Datta[1,2], J. D. Bowman[3,4], and C. L. Carilli[2]
[1] New Mexico Tech, Socorro, NM 87801, USA; adatta@nrao.edu
[2] National Radio Astronomy Observatory, Socorro, NM 87801, USA
[3] California Institute of Technology, Pasadena, CA 91125, USA
Received 2010 May 21; accepted 2010 September 17; published 2010 November 3

ABSTRACT

The H I 21 cm transition line is expected to be an important probe into the cosmic dark ages and epoch of reionization. Foreground source removal is one of the principal challenges for the detection of this signal. This paper investigates the extragalactic point source contamination and how accurately bright sources ($\gtrsim$1 Jy) must be removed in order to detect 21 cm emission with upcoming radio telescopes such as the Murchison Widefield Array. We consider the residual contamination in 21 cm maps and power spectra due to position errors in the sky model for bright sources, as well as frequency-independent calibration errors. We find that a source position accuracy of 0.1 arcsec will suffice for detection of the H I power spectrum. For calibration errors, 0.05% accuracy in antenna gain amplitude is required in order to detect the cosmic signal. Both sources of subtraction error produce residuals that are localized to small angular scales, $k_\perp \gtrsim 0.05\,\mathrm{Mpc}^{-1}$, in the two-dimensional power spectrum.

Key words: early universe – intergalactic medium – methods: data analysis – radio lines: general – techniques: interferometric

Online-only material: color figures

1. INTRODUCTION

The cosmological epoch of reionization (EoR) is a key milestone in the history of structure formation, marking the transition from a fully neutral to a highly ionized intergalactic medium (IGM) due to the ultraviolet and X-ray radiation of early stars, galaxies, and black holes. Recent observations of the Gunn–Peterson effect, i.e., Lyα absorption by the neutral IGM, toward the most distant quasars ($z \sim 6$), and the large scale polarization of the CMB, corresponding to Thompson scattering during reionization, have set the first constraints on the reionization process. These results suggest significant variance in both space and time, starting perhaps as far back as $z \sim 11$ (Komatsu et al. 2010; *WMAP* seven years data) and extending to $z \sim 6$ (Fan et al. 2006). Previous *WMAP* five years data indicate the 5σ detection of the *E*-mode of polarization which rules out any instantaneous reionization at $z \sim 6$ at 3.5σ level. In the case of the Gunn–Peterson effect, the IGM becomes optically thick to Lyα absorption for a neutral fraction as small as $\sim 10^{-3}$. In order to overcome these limitations, it has been widely recognized that mapping the redshifted H I 21 cm line has great potential for direct studies of the neutral IGM during reionization (Barkana & Loeb 2001; Fan et al. 2006; Furlanetto et al. 2006; Morales & Wyithe 2009).

There are a number of upcoming low-frequency arrays with key science goals to detect the H I 21 cm signal from the EoR. This includes the Murchison Widefield Array (MWA; Mitchell et al. 2008; Lonsdale et al. 2009; Bowman et al. 2006), Precision Array to Probe Epoch of Reionization (PAPER; Backer et al. 2007; Parsons et al. 2010), Low Frequency Array (LOFAR; Harker et al. 2010; Jelić et al. 2008; Labropoulos et al. 2009), and Giant Meterwave Radio Telescope (GMRT; Pen et al. 2009; Paciga et al. 2010). One of the major challenges for all of these upcoming arrays will be the removal of the continuum foreground sources in order to detect the faint H I signal from the EoR.

A variety of continuum foregrounds complicate redshifted 21 cm measurements of the EoR (Shaver et al. 1999). Diffuse Galactic synchrotron emission dominates the low-frequency radio sky and is approximately 4 orders of magnitude brighter than the $\sim$10 mK 21 cm signal at the frequency relevant to reionization ($\nu \approx 150$ MHz). In addition, Galactic and extragalactic free–free emission contribute additional flux to the diffuse foreground. Radio point sources from active galactic nuclei (AGNs), radio galaxies, and local Galactic sources are numerous and particularly challenging. The brightest of these sources have fluxes well above $S > 1$ Jy and are 7 or 8 orders of magnitude above the EoR signal in low-frequency radio maps. The distribution of point sources also extends to very faint levels such that the brightness temperature due to confused sources in upcoming arrays will be $\sim$10 K or 3 orders of magnitude brighter than the 21 cm background.

In this paper, we discuss how the radio interferometric imaging techniques are going to affect the foreground source modeling and subsequent removal from the data set in order to search for the EoR signal. Recently, there has been extensive research on foreground source modeling at these low frequencies (Di Matteo et al. 2002; Jelić et al. 2008; Thomas et al. 2009). Similar effort has also been made in exploring different techniques to remove the foregrounds from the EoR data set by Morales et al. (2006a, 2006b), Bowman et al. (2009), Liu et al. (2009), Parsons et al. (2010), Gleser et al. (2008), and Harker et al. (2009a, 2009b). Since attempting to observe a signal below the confusion limit of foreground sources is a novel aspect of 21 cm experiments, most of these works primarily focus on the removal of faint and confused sources that fall below a specified cutoff flux limit, S_{cut} ($\approx$1 Jy). They do not consider the foreground sources brighter than S_{cut} and how accurately they need to be removed. Indeed, most of these analyses implicitly assume that the bright foreground sources above S_{cut} have been removed perfectly. But in reality imperfect instrument calibration or any

[4] Hubble Fellow.

errors in the subtracted foreground model will introduce artifacts and leave residual contamination in the data after bright source removal, even by traditional techniques such as "peeling." These residuals may interfere with either the subsequent faint source subtraction or the ultimate detection and characterization of the redshifted 21 cm signal.

In Datta et al. (2009), we dealt with the bright point sources above $S_{\rm cut}$ and the limitations that will be caused due to imperfect removal of such sources in the image plane. In this paper, we extend the initial analysis in order to estimate the residual contamination in the power spectral domain of improper bright source subtraction. The objective of this paper is to demonstrate how the accuracy in the foreground removal affects the detection of H I 21 cm power spectra with the MWA. In Section 2, we discuss our choice of sky model and outline the simulation parameters, including the array specifications and data reduction procedure, and describe the two categories of corruption terms that we will consider: source position errors and residual calibration errors. The results obtained for the residual angular power spectrum, spherically averaged three-dimensional (3D) power spectrum, and two-dimensional (2D) power spectrum are presented in Section 3. Finally, in the last section, we summarize the implications of the results from our simulations.

2. THE SIMULATION

2.1. Sky Model

Our main aim is to explore the level of accuracy needed in instrument calibration and foreground modeling in order to ensure that the residual errors from bright foreground source removal do not obscure the detection of the signal from cosmic reionization. With this goal in mind, we use a simple sky model for our simulations that only includes bright radio point sources. No diffuse emission from the Galaxy is included as a part of the sky model, and the 21 cm signal and thermal noise are also omitted. Our sky model is derived from the log N–log S distribution of sources and is termed the "Global Sky Model" (GSM) from now onward. Since the GSM only includes sources above 1 Jy, we follow the source counts from the 6C survey at 151 MHz (Hales et al. 1988):

$$dN/dS = 3600 S_{\rm Jy}^{-2.5}\ {\rm Jy}^{-1}\ {\rm str}^{-1}. \quad (1)$$

For a field of view of 15° the total number of sources above 1 Jy is ∼170, following the above power-law distribution. The entire flux range, between 1 Jy and 10^3 Jy, has been subdivided into several bins (in the logarithmic scale) and populated with the number of sources that corresponds to the flux range of each bin (according to Equation (1)). Inside each bin, we have assigned each source a flux density following a normal distribution. The strongest source in our GSM is ∼200 Jy. The observed distribution of radio sources shows evidence for only very weak angular clustering and the brightest extragalactic sources in the sky are not clustered at all (Blake & Wall 2002). Therefore, in order to assign a position to each of these sources within the field of view, a uniform random number generator has been used which predicted the offset from the field center for respective sources. In the GSM all the foreground sources are flat spectrum, i.e., with zero spectral index ($\alpha = 0$).

Figure 1 shows a simulated image of our bright source sky model that has been produced using the procedure described below. A wide-field variant of the well-known Clark-CLEAN algorithm that utilizes a w-projection algorithm for 3D imaging was applied to the simulated image. The apparent angular size of the sources in Figure 1 reflects the size of the synthesized beam. The input sources in the model are treated as ideal point sources. This input sky model is used for all the simulations presented in this paper.

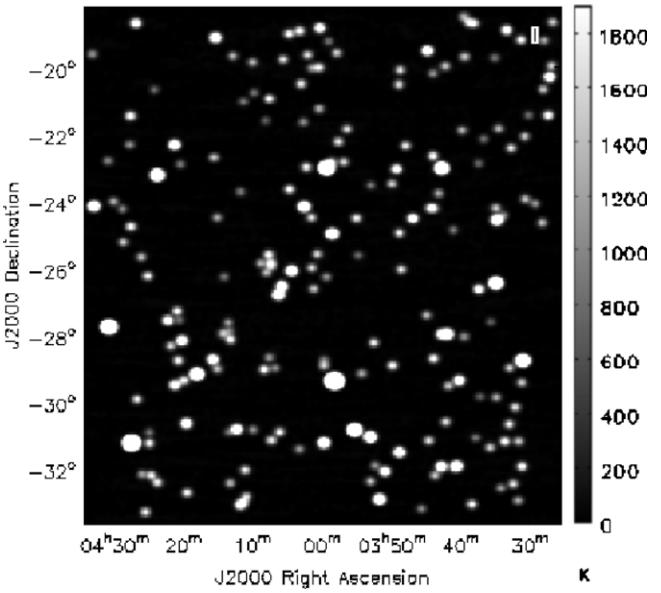

Figure 1. Simulation of the sky model centered on R.A. = 4 hr and decl. = −26° as would be observed by the MWA. Clark-CLEAN has been applied to this image using w-projection (256 planes) and natural weighting (Datta et al. 2009).

Table 1
Array Specifications

Parameters	Values
No. of tiles	512
Central frequency	158 MHz ($z \sim 8$)
Field of view	∼15° at 158 MHz ($\propto \lambda$)
Synthesized beam	∼4.′5 at 158 MHz ($\propto \lambda$)
Effective area per tile	∼17 m^2
Maximum baseline	∼1.5 km
Total bandwidth	32 MHz
$T_{\rm sys}$	∼250 K
Channel width	∼32 kHz
Thermal noise	∼7.55 mK
	(5000 hr and 2.5 MHz)

Notes. Array parameters have been influenced by the MWA specifications as mentioned in Mitchell et al. (2008) and Bowman et al. (2009). The actual MWA field of view is ∼25° at 150 MHz.

2.2. Array Specifications

Table 1 outlines the instrumental parameters that we have assumed for this analysis. Most of these parameters reflect the current specifications for the MWA, but we note that the array is presently under development and some properties may be subject to change. In addition, we have intentionally reduced the simulated field of view compared to the actual MWA in order to reduce the computational overhead of the simulation. Figure 2 shows the array layout for the 512 element array with maximum baseline of 1.5 km.

For the purposes of modeling earth rotation synthesis in the instrumental response, the center of the target field is

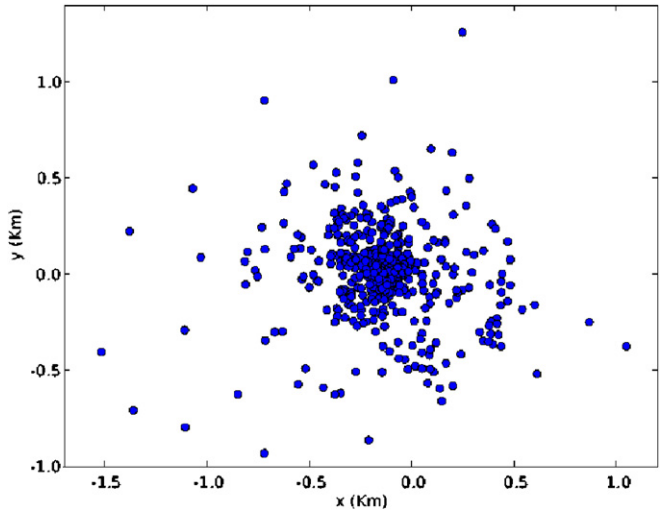

Figure 2. Array layout for the 512 elements with a maximum baseline of 1.5 km. (A color version of this figure is available in the online journal.)

chosen such that it coincides with one of the cold spots in the foreground Galactic synchrotron emission visible from the southern hemisphere location of the MWA. The exact field center used for the GSM is 4 hr in right ascension and -26° in declination. Most of the upcoming low-frequency telescopes, including the MWA, will only be able to observe a field around its transit. We have used 6 hr of integrations for all the simulations, assuming that the telescopes will observe a field between ± 3 hr in Hour Angle from transit.

2.3. Data Reduction Procedure

The 15° field of view will include $\sim$170 bright sources ($>$1 Jy). The individual flux densities of these foregrounds are $\sim 10^5$–10^7 times higher than the signal from cosmic reionization that these instruments are aiming to detect. So the challenge lies in calibration and subsequent removal of such bright sources from the raw data sets. The data rate of 19 GB s^{-1} (Mitchell et al. 2008) will not allow the MWA to store the raw visibilities produced by the correlator. Hence real-time calibration and imaging needs to be done in order to reduce the data volume and store the final product in the form of image cubes (Mitchell et al. 2008). The critical steps include removal of the bright sources above the $S_{\rm cut}$ level from the data sets in these iterative rounds of real-time calibration and imaging procedure. As a result the residual image cubes will not be dominated by these bright sources and the rest of the foregrounds can be removed in the image domain.

However, the accuracy of the foreground source removal strategies are strongly dependent on the data reduction procedure. The likely data reduction procedure which will be followed by the upcoming telescopes can be broadly outlined as

1. the raw data sets from the correlator will go through real-time calibration and subsequent removal of the bright sources based on some GSM, down to $S_{\rm cut}$ level, in the UV domain.
2. The residual data sets will be imaged and stored as a cube for the future processing and removal of sources which are below $S_{\rm cut}$.

The simulated data reduction pathway that we follow in this paper is

1. first, the observed visibilities ($V_{ij}^{\rm Obs}(\mathbf{u}, \nu) \equiv V_{ij}^{\rm GSM_{perfect}}(\mathbf{u}, \nu)$) are simulated for a 6 hr observation (± 3 hr in Hour Angle) using the GSM and the array configuration from Section 2.2. In the above notation, $\mathbf{u} \equiv (u, v)$ denotes the Fourier conjugate of the sky coordinate (θ_x, θ_y) and ν is the frequency of observations.
2. Next, we generate the foreground model ($V_{ij}^{\rm mod}(\mathbf{u}, \nu)$) that will be subtracted from the observation. In this case, the model is corrupted to either simulate errors in the assumed positions of the sources or to simulate calibration errors. For the source position errors, the model visibilities are given simply by

$$V_{ij}^{\rm mod}(\mathbf{u}, \nu) = V_{ij}^{\rm GSM_{imperfect}}(\mathbf{u}, \nu), \quad (2)$$

where the position of each source has been slightly moved from its original location by a distance drawn from a Gaussian distribution with standard deviation σ_θ. We assume that the source position errors are constant throughout the entire duration of the observation, as would be the case for a foreground model constructed from either an outside catalog or from the data itself at the conclusion of the observation. This is an idealization that may be broken in practice if sources are "peeled" in real time.
For the residual calibration errors, the model visibilities are given by

$$V_{ij}^{\rm mod}(\mathbf{u}, \nu) = g_i(t) g_j^*(t) V_{ij}^{\rm GSM_{perfect}}(\mathbf{u}, \nu), \quad (3)$$

where $g_i(t) \approx (1 + a_i) e^{i\phi_i}$ are the antenna-dependent complex gains. The parameters a_i and ϕ_i denote small amplitude and phase deviations, respectively, and are each drawn from their own Gaussian distribution with standard deviation σ_a or σ_ϕ (Datta et al. 2009).

The MWA will produce calibration solutions in real time with an $\sim$8 s cadence. This rapid pace is planned in order to simultaneously calibrate both the instrument hardware properties and the ionospheric phase screen. It is not known, yet, if in practice the residual calibration errors from the real-time processing will be largely independent or highly correlated between individual 8 s solutions. This is an important experimental property to consider in our simulation, because the degree of correlation greatly affects the level of accuracy needed in individual calibration solutions. If the individual errors are largely independent, then each 8 s sample can be modeled as coming from a Gaussian distribution and the accuracy tolerance will be relatively loose since many samples will be available and tend to average toward zero. Such a situation would be the best-case scenario. On the other hand, if the calibration errors are highly correlated, then each calibration solution must meet a much more stringent accuracy level to achieve the same residual contamination at the end of the full integration.

For our simulation procedure, we assume a relatively conservative scenario that the residual errors in a given antenna's 8 s calibration solutions are perfectly correlated for the duration of one 6 hr observing night, but perfectly uncorrelated between successive observing nights. We further assume that the residual errors *between* antennas are perfectly uncorrelated at all times. This choice is somewhat arbitrary given the current level of knowledge, but we believe it is a plausible fiducial case since both the

overall ionospheric properties and the ambient conditions may change significantly from day-to-day. Hence, in our simulation, σ_a and σ_ϕ are used to draw a calibration error value (a_i and ϕ_i) from a Gaussian distribution only once per antenna per night and that specific error is applied to all the simulated 8 s solutions for the given antenna throughout the 6 hr period of rotation synthesis. When the next night's observing block commences, a new error is drawn from the distribution for each antenna, and so on.

3. Now we are ready to calculate the residual visibilities by subtracting the foreground model produced in step (2) from the simulated observation of step (1) according to

$$V_{ij}^{\rm res}(\mathbf{u}, \nu) = V_{ij}^{\rm obs}(\mathbf{u}, \nu) - V_{ij}^{\rm mod}(\mathbf{u}, \nu). \quad (4)$$

In the sections below, we refer to this step as GSM subtraction and it was implemented using the UVSUB algorithm (Cornwell et al. 1992). For the residual calibration errors, we can reduce Equation (4) by substituting in with Equation (3) and simplifying to obtain

$$V_{ij}^{\rm res}(\mathbf{u}, \nu) = (1 - g_i g_j^*) V_{ij}^{\rm GSM_{perfect}}. \quad (5)$$

4. At this point, we have completed the subtraction of bright sources from our simulated observation, leaving only residual contamination due to the differences between our simulated observation and the corrupted model. Example images of the residual contamination at this stage are shown in Datta et al. (2009, panels (a) of their Figures 5, 7, and 9). In practice, this bright source-subtracted data cube will be the starting point for the second stage of redshifted 21 cm foreground subtraction that aims to remove faint and confused sources by fitting and subtracting a low-order polynomial along the frequency axis for each line of sight in the data cube. We want to understand how this additional process affects the end result of the bright source removal, so we approximate the faint source polynomial fitting here by applying a Fourier transform to the UV map generated from the residual visibilities in order to produce a residual dirty image cube, $I^{\rm res}$. In this dirty image cube, we fit a third-order polynomial in frequency along each line of sight and subtract it. Thus, we obtain the final residual image, $I_{\rm polysub}^{\rm res}(\vec{\theta}, \nu)$. We refer to this step as GSM+polynomial subtraction for the remainder of this paper and it was implemented using the IMLIN algorithm (Cornwell et al. 1992). To illustrate this final result, Figure 3 shows residual spectral profiles along four lines of sight in the dirty image cube after polynomial fitting and subtraction. Example images of the final residual contamination after GSM+polynomial subtraction are also shown in Datta et al. (2009, panels (b) of their Figures 5, 7, and 9).

Using higher order polynomials in the GSM+polynomial subtraction step removes structures at increasingly smaller scales (McQuinn et al. 2006). This improves the foreground cleaning, but since the 21 cm reionization signal has significant structures on scales that correspond to ∼2.5 MHz or approximately 10% of the bandwidth over which the polynomial is fit, it also has the potential to remove much of the 21 cm signal. We have restricted our attention to a third-order polynomial in this work because it is the lowest-order polynomial likely to be sufficient for removing the faint continuum sources given their power-law spectral shapes.

We also explored using the UVLIN algorithm (Cornwell et al. 1992), which fits and subtracts polynomials in the UV domain instead of the image domain, eliminating the need to convert our residual data sets into image cubes. However, UVLIN works perfectly only within a small field of view, depending on the channel width in frequency (Cornwell et al. 1992), and was found to be inadequate for our purposes.

5. The final step in our procedure is to calculate power spectra from the residual image cubes and compare these residual foreground power spectra to the theoretically predicted 21 cm power spectrum and expected thermal noise power spectrum for the MWA. We calculate three forms of the residual power spectra from our final data cubes: the derived angular power spectrum C_ℓ for a narrow frequency channel, the spherically averaged 3D power spectrum $P(k)$ from the entire data cube, and the 2D power spectrum $P(k_\perp, k_\parallel)$ found by averaging over transverse modes in the full 3D power spectrum. Each of these cases is discussed in more detail in Section 3. As a reference, we show in Figure 4 the spherically averaged power spectrum for our input GSM before any source removal has been applied.

In order to simulate the observed visibilities ($V_{ij}^{\rm Obs}$), we have used the simulator tool in the CASA software.[5] We have also used CASA to perform the imaging, GSM subtraction, and the subsequent GSM+polynomial subtraction step. The rest of the operations are performed using separately written python[6] scripts.

3. RESULTS

We begin our discussion of the results of the residual power spectrum determination by reviewing our initial findings from Datta et al. (2009). In that work, we explored the source position and calibration accuracy needed to allow direct imaging of Strömgren spheres with very deep integrations by the MWA. Our simulations demonstrated that knowledge of the true positions of the bright foreground sources in an MWA target field is required to within $\sigma_\theta = 0.1$ arcsec, assuming Gaussian errors, in order for the residual contamination following subtraction to be below the 21 cm signal from Strömgren spheres in image maps that could be acquired by the MWA with 5000 hr of integration. Similarly, in Datta et al. (2009) we found that, for the case of calibration errors corrupting the measurements under the same conservative assumptions outlined in step (2), a calibration accuracy of $\sigma_a = 0.2\%$ systematic error in gain amplitude per night of observing (or $\sigma_\phi = 0.2^\circ$ in phase) is needed for the residual contamination to be below the thermal noise in a part of the image map far from any bright sources for a long integration by the MWA.

3.1. Motivation for Power Spectra

In this work, we focus our attention on the residual contamination that can be tolerated in measurements of 21 cm power spectra, rather than direct imaging of the 21 cm background. The 21 cm field from the reionization epoch, as observed by a radio interferometer, is a fully 3D data set, where the two angular dimensions translate to transverse distances on the plane of the sky and the spectral dimension corresponds to a line-of-sight distance within the target field through the redshift of the 21 cm line (modulo small deviations due to peculiar velocities). The

[5] http://casa.nrao.edu/
[6] http://www.python.org/

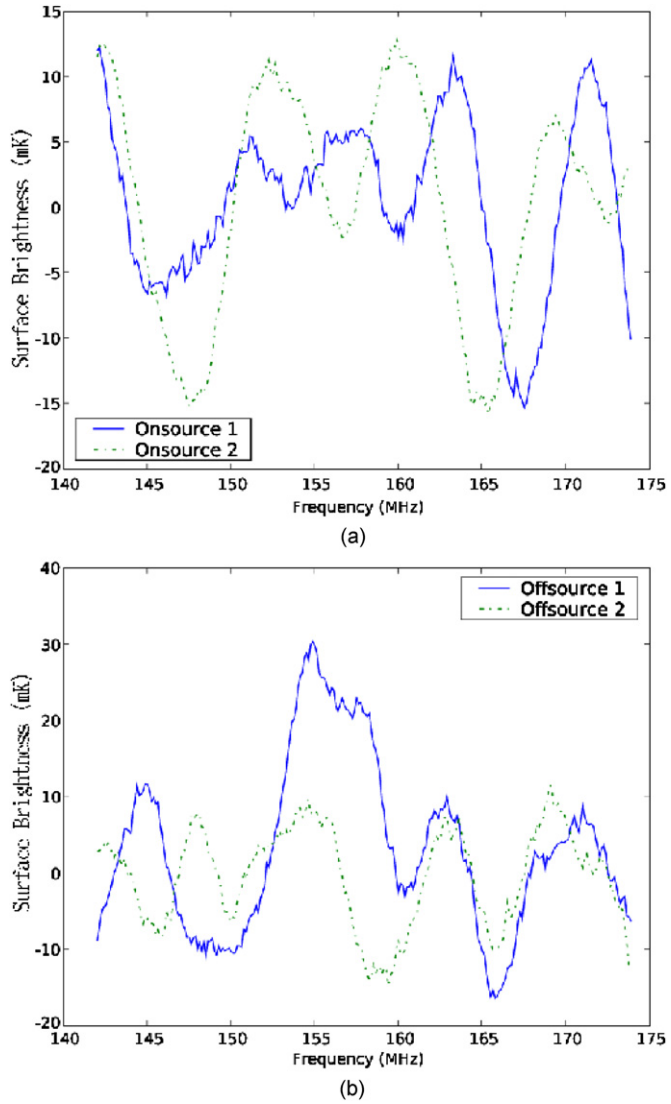

Figure 3. (a) Spectral profile along two lines of sight in the final residual image cube following the GSM+polynomial subtraction step. In this case, the two pixels were chosen to be next to the positions of sources in the input model. (b) Same as panel (a), but here, the two pixels were chosen to be far from any sources in the input model. Synthesized beam area of 4.5 arcmin × 4.5 arcmin is used to convert the flux densities to surface brightness.

(A color version of this figure is available in the online journal.)

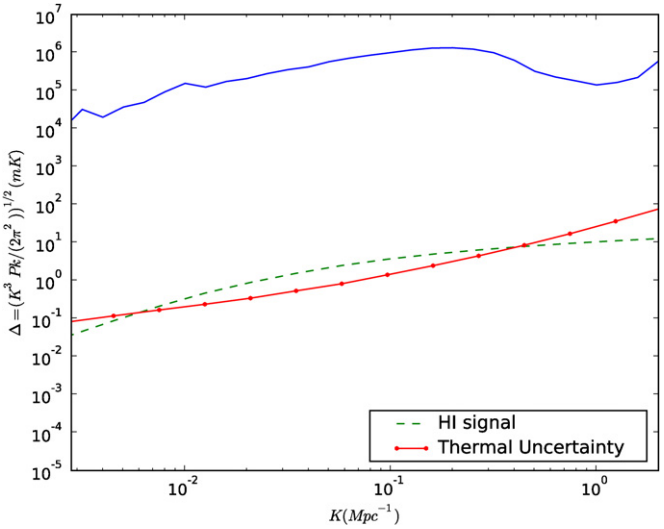

Figure 4. 1D spherically averaged power spectrum of the input GSM showing the total power of the bright sources in the sky model before any foreground removal has been applied. The thermal noise uncertainty for a 300 hr observation by the MWA is also shown, along with the H I 21 cm signal power spectrum for a fully neutral IGM ($x_{\rm H\,I} = 1$) at $z = 8$ (Furlanetto et al. 2006).

(A color version of this figure is available in the online journal.)

full treatment of the power spectrum of such an observation is also 3D.

The primary motivation for seeking to first detect and characterize 21 cm power spectra, rather than immediately attempt to image the background, is that the signal-to-noise ratio (S/N) in a properly chosen power spectrum measurement can be significantly higher than the per-pixel S/N in a direct imaging observation. This is because the information in the 21 cm field is expected to be effectively compressed in a power spectrum by averaging over many Fourier modes.

As we briefly mentioned in step (5) above (Section 2.3), we consider three different power spectra that can be calculated from MWA observations. The first is the angular power spectrum, found by taking the 2D Fourier transform of an image map at any specific frequency in the observed data cube. This type of power spectrum is widely used for characterizing CMB anisotropies and was the first to be proposed in connection to redshifted 21 cm measurements. For 21 cm measurements, however, it has become recognized over the last several years that angular power spectra will be highly susceptible to contamination by diffuse foregrounds (Di Matteo et al. 2002, 2004) and will likely have limited utility for characterizing 21 cm fluctuations. Nevertheless, most existing experiments, including PAPER (Parsons et al. 2010) and the GMRT EOR project (Paciga et al. 2010), use angular power spectra as effective diagnostics of their instrumental performance and foreground subtraction routines. We begin our investigation with angular power spectra in Section 3.2.

More recently, the preferred method of deriving a power spectrum for characterizing redshifted 21 cm fluctuations has been to average over the full 3D 21 cm power spectrum, $P(\mathbf{k})$, in logarithmic shells of constant radius, k, thus yielding $P(k)$ (Morales & Hewitt 2004). This method is possible because the 21 cm field is homogeneous and isotropic, and hence, will have spherical symmetry in the 3D power spectrum (in practice, redshift-space effects due to peculiar velocities distort this symmetry, however; albeit in a reasonably predictable manner). The spherically averaged power spectrum has the advantage of greatly increasing the S/N over the angular power spectrum by effectively reducing the entire observed data set into ∼10 independent measurements (as opposed to reducing only a single image plane in the case of the angular power spectrum). For the MWA, this means that only ∼300 hr of observing are required to have sufficient sensitivity to detect the spherically averaged 21 cm power spectrum at $z \approx 8$, assuming the IGM is not fully ionized at that time, whereas ∼5000 hr would be necessary to directly image the 21 cm background. We present results for the spherically averaged power spectrum of an entire simulated MWA data cube in Section 3.3.

Finally, in Section 3.4, we consider an intermediate level between the spherically averaged power spectrum and the full 3D power spectrum. In this case, we hope to facilitate an approach that retains the S/N benefits of the spherically averaged power spectrum, while anticipating some of the possible problems. It has been shown (Bowman et al. 2009) that different regions in the full 3D power spectrum will be corrupted

by diffuse foregrounds to varying levels. But in our somewhat "naive" spherically averaged treatment described above, we assume that the entire 3D power spectrum derived from MWA observations contains equally valid measurements. It would be better if we could assess which regions of the full 3D power spectrum to use before performing the average and then discard the regions anticipated to be most corrupted by foreground contamination when averaging. This should be possible with a 2D power spectrum that is calculated in a manner similar to the spherically averaged power spectrum, but, instead of averaging over $P(\mathbf{k})$ in shells of constant k, we average over transverse annuli of constant $k_\perp \equiv \sqrt{k_x^2 + k_y^2}$ for each line-of-sight plane $k_\parallel \equiv k_z$ in the 3D power spectrum. This yields the 2D power spectrum, $P(k_\perp, k_\parallel)$. Because both the diffuse and bright source foregrounds are generally isotropic in the sky, they should exhibit axial symmetry around the line-of-sight axis in the 3D power spectrum, as opposed to the (approximate) spherical symmetry of the 21 cm signal (Morales et al. 2006b). Hence, it should be possible to retain sufficient information to largely isolate the foreground and signal contributions in $P(k_\perp, k_\parallel)$. We anticipate that the 2D power spectrum will likely be the default technique applied to most 21 cm measurements as experiments progress, hence the results in Section 3.4 should be the most pertinent to many readers.

Because power spectra measurements differ significantly from direct imaging, there are several key questions that we seek to address in the remainder of this section: (1) are the tolerances on the source position and calibration errors greater (or lesser) than in the direct imaging case, (2) is a particular region of the power spectrum likely to be more affected by residual contamination than another, and (3) can we hope to build a library of template models for foreground contamination that could be used to marginalize out some of the contamination during the analysis of the power spectrum?

We will address these questions for each of the three classes of power spectra listed above. For each class of power spectrum, we will present residual power spectra for both corruption models: source position errors and calibration errors. And for each corruption model, we will use three fiducial levels of error in our investigation: for source position errors, our fiducial cases are $\sigma_\theta = \{0.01, 0.1, \text{and } 1 \text{ arcsec}\}$ while for the calibration errors our fiducial levels are $\sigma_a = \{0.01\%, 0.1\%, \text{and } 1\%\}$ in gain amplitudes, which also translates to $\sigma_p = \{0.^\circ01, 0.^\circ1, \text{and } 1^\circ\}$ in phase (Datta et al. 2009).

3.2. Angular Power Spectrum

White et al. (1999) describe the technique to derive the angular power spectrum from radio interferometric data. Using the flat-field approximation (Datta et al. 2007)

$$C_\ell = \frac{\sum_{2\pi|\mathbf{u}|=\ell} W(\mathbf{u})|V(\mathbf{u})|^2}{\sum_{2\pi|\mathbf{u}|=\ell} W(\mathbf{u})}, \quad (6)$$

where $|\mathbf{u}| = \sqrt{u^2 + v^2}$ and $\ell \simeq 2\pi|\mathbf{u}|$ under flat-field approximation. Here, $V(\mathbf{u}, \nu)$ is the un-weighted visibilities from the residual images and $W(\mathbf{u})$ is the number of visibilities entering each $\mathbf{u}$ cell.

In Figures 5 and 6, we have plotted $\ell(\ell+1)C_\ell/(2\pi)$ calculated for one frequency bin of width 125 KHz from our simulated residual image maps. Figure 5 shows the angular power spectrum resulting from using the foreground model that is corrupted by source position errors. Figure 6 illustrates the same result for the case of residual calibration errors. The top panels (a) of both figures show the angular power spectra derived from the GSM-subtracted residuals of step (3) in our analysis procedure. The bottom panels (b) show the angular power spectra from the final GSM+polynomial-subtracted residuals following step (4). The shaded region in panel (b) of both figures corresponds to $\ell \leqslant 250$. The GSM+polynomial subtraction step, which involves fitting a third-order polynomial over a total bandwidth of 32 MHz, is expected to remove most of the significant structures for scales larger than this. All of the plots have been restricted to $\ell \lesssim 5000$ to match the size of the MWA synthesized beam (4.5 arcmin).

The total thermal noise power is much stronger than the angular power spectra of the H I 21 cm signal. Hence, we have

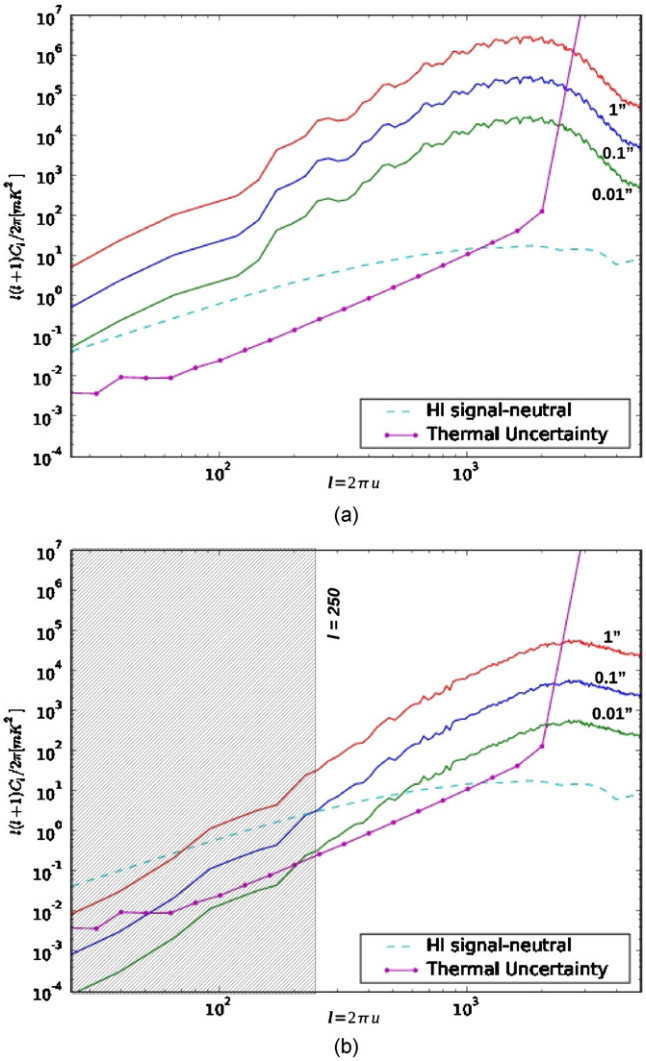

Figure 5. (a) Angular power spectrum of the GSM-subtracted residual image $I^{\rm res}(\vec{\theta}, \nu)$ made after subtraction of a foreground model with source position errors of $\sigma_\theta = \{0.01, 0.1, \text{and } 1 \text{ arcsec}\}$. (b) Same as panel (a) but for the GSM+polynomial-subtracted residual image, $I^{\rm res}_{\rm polysub}(\vec{\theta}, \nu)$, that is produced after polynomial fitting and subtraction has been applied along each sight line. The shaded region in panel (b) corresponds to $\ell \leqslant 250$, where the polynomial fitting is expected to remove most of the structure. Both panels include the thermal noise uncertainty power spectrum assuming 5000 hr of observation with the MWA and the H I 21 cm signal power spectrum for a fully neutral IGM ($z \sim 8$, $x_{\rm H\,I} = 1$; Furlanetto et al. 2006). These angular power spectra are what would be expected from the MWA if it integrated deep enough to directly image a typical cosmic Strömgren sphere.

(A color version of this figure is available in the online journal.)

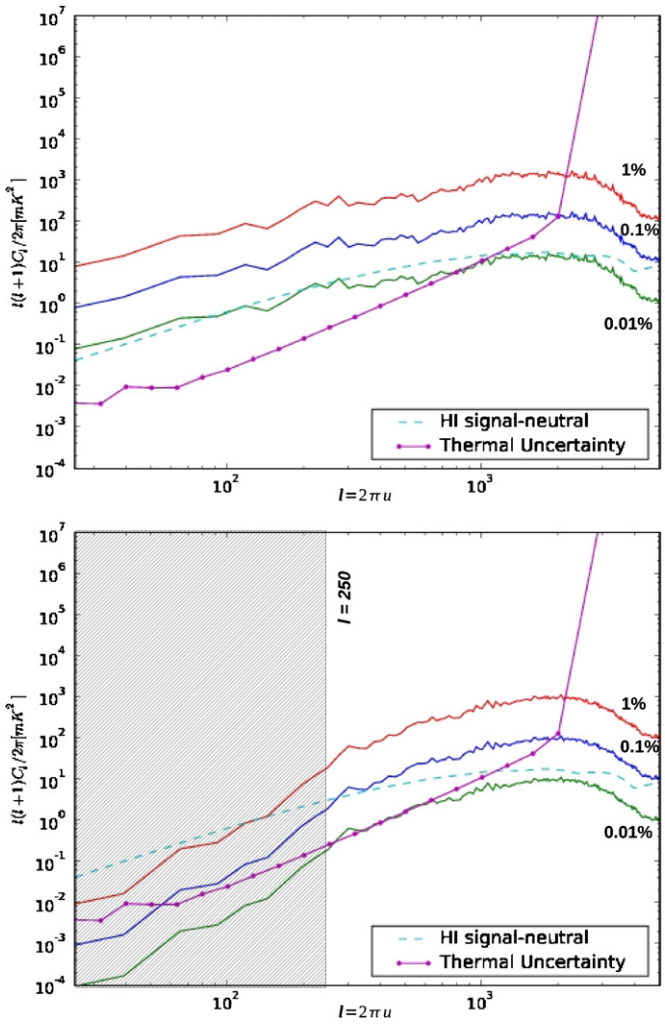

Figure 6. Same as Figure 5, but for the residuals due to calibration errors. In this case, the three residual angular power spectra are for errors of $\sigma_a = \{0.01\%, 0.1\%, \text{and } 1\%\}$.

(A color version of this figure is available in the online journal.)

assumed that the final power spectrum from the real data will be generated by dividing the observation into different epochs of equal duration and then cross-correlating the data cubes from the two epochs (Bowman et al. 2009). This approach preserves the persistent H I 21 cm signal and eliminates the thermal noise power (which will be independent between the two observing epochs and, therefore, average to zero during the cross-correlation), leaving only the thermal uncertainty. Hence, the relevant noise figure for the angular power spectra measurement is given by

$$C_\ell^N = \left\langle \frac{\sum_\ell |N_1(\ell)^* N_2(\ell) + N_2(\ell)^* N_1(\ell)|}{2} \right\rangle, \quad (7)$$

where N_1 and N_2 are simulated noise measurements from two different epochs (Bowman et al. 2009).

The residual angular power spectra in the figures can be compared to the thermal noise *uncertainty* in the observations (Equation 7) and a fiducial 21 cm signal. We plot the expected thermal noise *uncertainty* angular power spectrum of the MWA after 5000 hr of integration assuming a system temperature of $T_{\rm sys} = 250$ K, channel width of 125 KHz, and the observing strategy described in Section 2.2. The thermal uncertainty spectrum is shown assuming the angular power spectrum has been binned in logarithmic intervals of width $\Delta\ell = 0.1$ or approximately 10 bins per decade. For the reference H I 21 cm signal, we show the power spectrum for a fully neutral IGM at $z = 8$ (Furlanetto et al. 2006). Modeling the 21 cm signal using a fully neutral IGM provides a reasonable fiducial expectation since recent reionization simulations (Lidz et al. 2008) show that the amplitude of the power spectrum over the scales probed by the MWA is likely to be even larger than the fully neutral level when the universe if roughly 50% ionized. It should be noted that different models predict different amplitudes for the H I power spectrum. For simplification, we have used this single realistic model to compare with our residual power spectrum. The specific conclusions regarding the scale-size dependence of where residual power will dominate the 21 cm signal will change depending on the reionization model.

Figure 5(a) shows that the angular power spectrum from the GSM-subtracted images are well above the thermal uncertainty power spectrum, as well as the model H I 21 cm signal power spectrum. In Figure 5(b), it is evident that the residual angular power in the GSM+polynomial-subtracted image is greatly reduced; and for two of our fiducial source position error levels ($\sigma_\theta = 0.1$ and 0.01 arcsec), the angular power spectra intercepts the H I signal power spectrum at $\ell \lesssim 700$. This shows that the GSM+polynomial-subtracted step is very crucial not only for removing faint and confused continuum foreground sources, but also for removing residual power left over after subtracting the bright foreground sources. A source position accuracy of $\lesssim$0.1 arcsec would allow the detection of H I 21 cm signal at $250 \lesssim \ell \lesssim 600$ scales.

Similar features are seen in Figure 6 for the case of calibration errors, where the residual angular power spectrum from the GSM-subtracted image only intercepts the thermal noise power spectrum near $\ell \sim 2000$ and only the $\sigma_a = 0.01$ % crosses below the model H I 21 cm signal power spectrum and the thermal uncertainty spectrum. Again, from Figure 6(b), it is evident that the residual angular power spectrum from the GSM+polynomial-subtracted image is much lower, particularly below the $\ell = 250$ threshold. We have not investigated in detail how scales larger than this threshold will be affected by the polynomial subtraction, but it is likely that some of the signal will be removed, as well. A calibration accuracy of $\sigma_a \lesssim 0.05\%$ should allow the detection of the H I 21 cm signal.

3.3. One-dimensional Spherically Averaged Power Spectrum

The spherically averaged 3D 21 cm power spectrum is the primary reionization observable targeted by the MWA. There has been extensive research on the statistical EoR power spectrum measurement of the brightness temperature fluctuations in low-frequency, wide-field radio observations. Detailed formulation has been developed in the literature by Morales & Hewitt (2004) and Zaldarriaga et al. (2004). The approaches described in these efforts are inspired by the techniques that have been employed successfully for interferometric measurements of CMB anisotropies (White et al. 1999; Hobson & Maisinger 2002; Myers et al. 2003). The primary approach is to convert the full 3D measurement cube to a one-dimensional (1D) power spectrum.

The first step is to transform our residual image cubes $I(\vec{\theta}, \nu)$ into $V(\mathbf{u}, \eta)$ by performing a 3D Fourier transform denoted by the operator $\mathbf{F}(\{\mathbf{u}, \eta\}, \{\vec{\theta}, \nu\})$. It should be noted here that before performing the Fourier transform, we have changed the units of the residual images from flux unit (Jy beam^{-1}) to brightness

temperature unit (mK). Hence, we get

$$V(\mathbf{u}, \eta) = \mathbf{F}(\{\mathbf{u}, \eta\}, \{\vec{\theta}, \nu\}) I(\vec{\theta}, \nu), \quad (8)$$

where $\mathbf{u}, \eta \equiv (u, v, \eta)$. Then, we transform the measurement coordinates $\mathbf{u}, \nu$ into the cosmological coordinates $\mathbf{k}$:

$$V(\mathbf{k}) = \mathbf{J}(\mathbf{k}, \{\mathbf{u}, \eta\}) V(\mathbf{u}, \eta) \quad (9)$$

$$= \mathbf{J}(\mathbf{k}, \{\mathbf{u}, \eta\}) \mathbf{F}(\{\mathbf{u}, \eta\}, \{\vec{\theta}, \nu\}) I(\vec{\theta}, \nu), \quad (10)$$

where $\mathbf{J}(\mathbf{k}, \{\mathbf{u}, \eta\})$ denotes the Jacobian of the coordinate transformation from $\mathbf{u}, \eta$ (in units of λ and Hz^{-1}) to $\mathbf{k}$ (in units of Mpc^{-1}). We have mainly followed the definition in Peebles (1993) and the formulation detailed in Morales & Hewitt (2004). Hence, we transformed a residual image cube (in sky coordinates) to a 3D residual visibility cube in the Fourier conjugate coordinates of comoving Mpc.

Assuming isotropy of space and ignoring redshift-space distortions inherent in converting our observed data cube to cosmological coordinates, the power spectrum can be taken as approximately spherically symmetric in cosmological $\mathbf{k} \equiv (k_x, k_y, k_z)$ coordinates. Hence, the power spectrum can be approximated to the square of the $V(\mathbf{k})$, averaged over spherical shells:

$$P(k) = \langle |V(\mathbf{k})|^2 \rangle_{|\mathbf{k}|=k}. \quad (11)$$

Thus, we obtain the 1D total power spectrum (Morales & Hewitt 2004) or the more common dimensionless power spectrum given by $\Delta^2 = k^3 P(k)/(2\pi^2)$.

While deriving the 1D power spectrum, we have weighted the individual measurements $|V(\mathbf{k})|^2$ by the per cell visibility contributions. This scheme is similar to the natural weighting scheme which is applied to the raw visibilities before imaging, and follows the form

$$P(k) = \frac{\sum_{|\mathbf{k}|=k} W_u(\mathbf{k}) |V(\mathbf{k})|^2}{\sum_{|\mathbf{k}|=k} W_u(\mathbf{k})}, \quad (12)$$

where $W_u(\mathbf{k})$ denotes the total number of visibilities contributing per $\mathbf{k}$ cell. Here, we should explicitly mention that the $V(\mathbf{k})$ used in the above equation are the un-weighted visibilities obtained from the residual images.

The total thermal noise power is much stronger than the 1D spherically averaged power spectra of the H I 21 cm signal. Hence, similar to the angular power spectrum case, we compare our results with the thermal noise uncertainty given by

$$P^N(k) = \left\langle \frac{\sum_{|\mathbf{k}|=k} |N_1(\mathbf{k})^* N_2(\mathbf{k}) + N_2(\mathbf{k})^* N_1(\mathbf{k})|}{2} \right\rangle, \quad (13)$$

where N_1 and N_2 are simulated noise measurements from two different epochs (Bowman et al. 2009).

Figures 7 and 8 show the 1D spherically averaged power spectrum from the residual images. As with the angular power spectrum, these figures also show theoretical H I 21 cm power spectrum. However, here, instead of using a total thermal noise power spectrum as we did for the angular power spectrum plots, we show the spherically averaged thermal noise *uncertainty* power spectrum from 300 hr of observation with the MWA, as mentioned in Equation 13. The thermal uncertainty spectrum is shown assuming that the spherically averaged power spectrum has been binned in logarithmic shells of width $\Delta k/k = 0.5$ or approximately five bins per decade. As discussed in Lidz et al. (2008), the MWA-512 will be sensitive primarily to scales

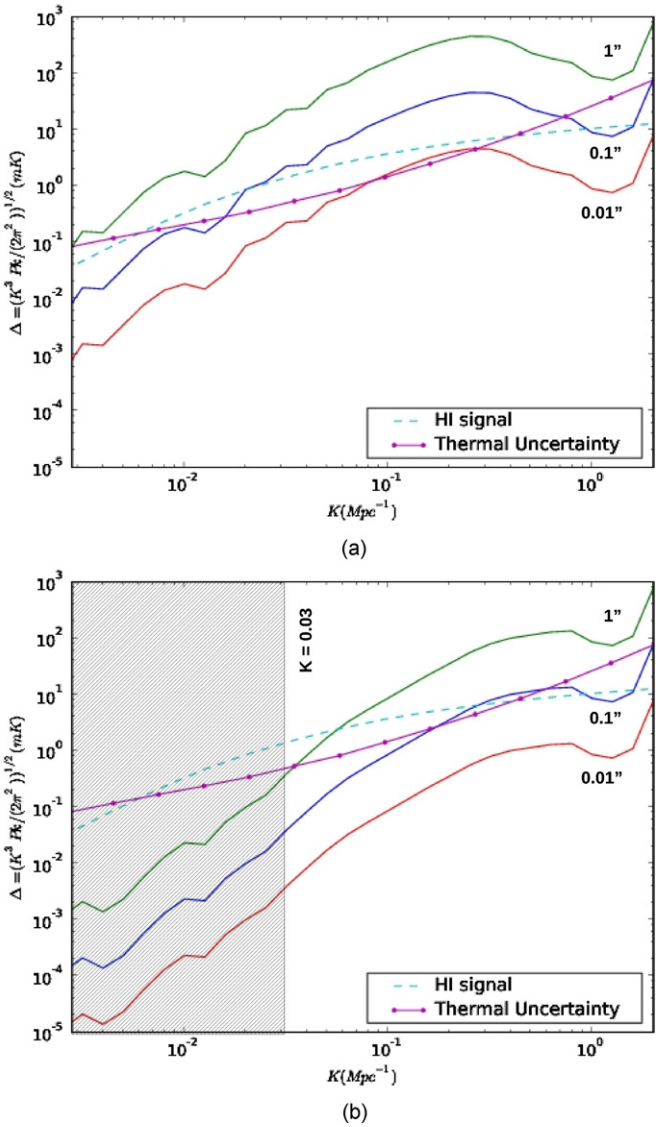

Figure 7. (a) 1D spherically averaged power spectrum of the GSM-subtracted residual image $I^{\mathrm{res}}(\vec{\theta}, \nu)$ made after subtraction of a foreground model with source position errors of $\sigma_\theta = \{0.01, 0.1, \text{and } 1 \text{ arcsec}\}$. (b) Same as panel (a) but for the GSM+polynomial-subtracted residual image, $I^{\mathrm{res}}_{\mathrm{polysub}}(\vec{\theta}, \nu)$, that is produced after polynomial fitting and subtraction has been applied along each sight line. The shaded region in panel (b) corresponds to $k \leqslant 0.03\,\mathrm{Mpc}^{-1}$, where the polynomial fitting is expected to remove much of the structure. Both panels include the thermal noise uncertainty spectrum assuming 300 hr of observation with the MWA and binning into logarithmic spherical shells of width $\Delta k/k = 0.5$ or approximately five bins per decade. The H I 21 cm signal power spectrum for a fully neutral IGM ($z \approx 8$, $x_{\mathrm{H\,I}} = 1$; Furlanetto et al. 2006) is also shown. Detecting the spherically averaged 21 cm power spectrum is the primary goal of the MWA.

(A color version of this figure is available in the online journal.)

$0.1 \lesssim k \lesssim 1\,\mathrm{Mpc}^{-1}$. The shaded regions in Figures 7(b) and 8(b) are at $k \leqslant 0.03\,\mathrm{Mpc}^{-1}$ and indicate the scales where the GSM+polynomial subtraction step removes significant power. These regions correspond to the $\ell \leqslant 250$ threshold in the angular power spectra plots. The higher end of the k value for the residual power spectrum is restricted due to the cell size in the image domain and frequency resolution of the channels in the residual image cube. The maximum value of $\mathbf{k}$ is attained along the k_z axis only, and hence few or no transverse (angular) modes contribute to the power spectrum at small scales above $k \gtrsim 0.6\,\mathrm{Mpc}^{-1}$ in Figures 7 and 8. The angular resolution of

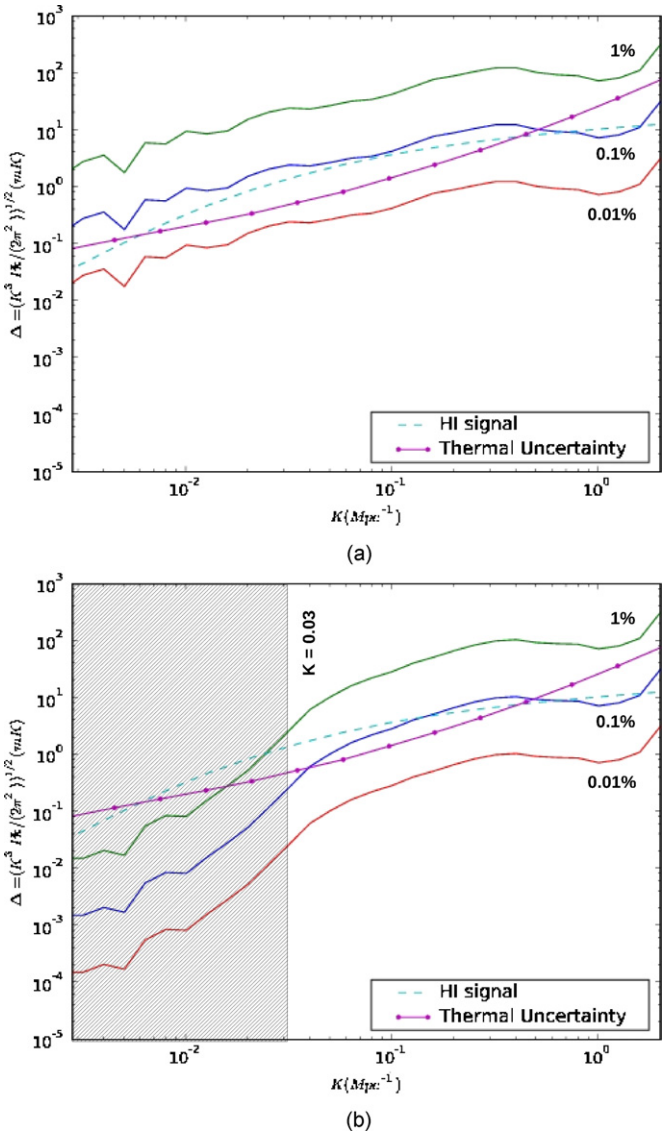

Figure 8. Same as Figure 7, but for the residuals due to calibration errors. In this case, the three residual spherically averaged power spectra are for errors of $\sigma_a = \{0.01\%, 0.1\%, \text{and } 1\%\}$.

(A color version of this figure is available in the online journal.)

the MWA of ∼4.5 arcmin (synthesized beam) corresponds to $k \approx 0.6\,\mathrm{Mpc}^{-1}$.

Figure 7 shows the 1D spherically averaged power spectra from the residual images. These are the residual images after the foreground subtraction in the presence of source position errors. Figure 7(a) shows that the 1D power spectra from the GSM-subtracted image with $\sigma_\theta = 0.01$ arcsec are only below the thermal uncertainty power spectrum and the H I signal power spectrum. In Figure 7(b), it is evident that the residual 1D spherically averaged power spectra with $\sigma_\theta = 0.01$ and 0.1 arcsec from the final GSM+polynomial-subtracted image are below the thermal uncertainty power spectrum and the H I signal power spectrum. Hence, a source position accuracy of $\sigma_\theta \lesssim 0.1$ arcsec would allow the detection of H I 21 cm signal with the MWA.

Turning to the case of the calibration errors, Figure 8(a) shows that only the 1D power spectrum from the GSM-subtracted image for calibration error of 0.01% is well below the thermal uncertainty power spectrum and the H I signal power spectrum. In Figure 8(b), it is evident that the 1D spherically averaged power spectra with $\sigma_a = 0.01\%$ and 0.1% from the GSM+polynomial-subtracted images are below the thermal uncertainty power spectrum and the theoretical H I signal power spectrum. Hence, the residual calibration accuracy of $\sigma_a \lesssim 0.05\%$ would allow the detection of H I 21 cm signal.

In comparison to the angular power spectrum, we can infer that the 1D spherically averaged power spectra have a better tolerance for both the source position and residual calibration errors. This also reflects the fact that the angular power spectrum has been produced using a single channel map of 125 KHz, whereas the 1D spherically averaged spectrum is produced with the total bandwidth of 32 MHz.

3.4. Two-dimensional Power Spectrum

In the previous section, we showed the analysis of the 1D spherically averaged power spectrum. However, this formulation mixes the contribution from the $k_\perp$ and $k_\parallel$ directions. It is useful, therefore, to break the averaging from the 3D **k**-space to the 1D k-space into two steps since both the foregrounds and a full treatment of the predicted redshifted 21 cm signal that includes redshifted-space distortions have *aspherical* structure in the Fourier domain (Morales et al. 2006b). Following McQuinn et al. (2006), we average over the transverse (angular) direction in the full three-dimension power spectrum to obtain $P(k_\perp, k_\parallel)$. We obtain the 2D power spectrum based on the maximum likelihood formalism following the same approach as used for the spherically averaged power spectrum in Equations (11) and (12).

Figures 9–12 illustrate the results of the simulation for the 2D residual power spectra. We show in Figure 9(a) the estimated 2D thermal noise uncertainty after 300 hr of integration with the MWA on our target field. Figure 9(b) shows a theoretical 2D H I signal power spectrum, $P(k_\perp, k_\parallel)$, of the H I signal in units of $\mathrm{mK}^2\,\mathrm{Mpc}^{-3}$. Figures 10 and 11 show the 2D power spectra of the residual image cubes for source position errors and calibration errors, respectively. In this Section, we have analyzed residual images for only one of our fiducial error levels for each type of model corruption. For source position errors, we use $\sigma_\theta = 0.1$ arcsec and for residual calibration errors, we use $\sigma_a = 0.1\%$.

3.4.1. The Wedge

Both types of subtraction errors we modeled resulted in a common, prominent wedge-like residual feature localized in the high-$k_\perp$, low-$k_\parallel$ region of the 2D power spectrum. The overall shape of the wedge-like feature is due to the 2D power spectrum of the point-spread function of the MWA (Figure 12), found by weighting the UV plane according to the number-density of visibilities at each sampled frequency and transforming into the $P(k_\perp, k_\parallel)$ coordinate frame. One can visualize the origin of such a wedge-like feature as the 2D power spectrum of a dirty map data cube from the MWA that consists of a single point source.

This basic underlying cause of the wedge feature illustrates the challenges inherent in bright source subtraction. As long as the MWA—or any radio interferometer—produces integrated maps in the "optimal" way using natural weighting of the visibilities, this feature (or a comparable one for a different instrument) will be a common by-product from point-like sources in the target field.

The secondary peak in residual power running parallel to the wedge-like feature at approximately twice the $k_\parallel$ value of

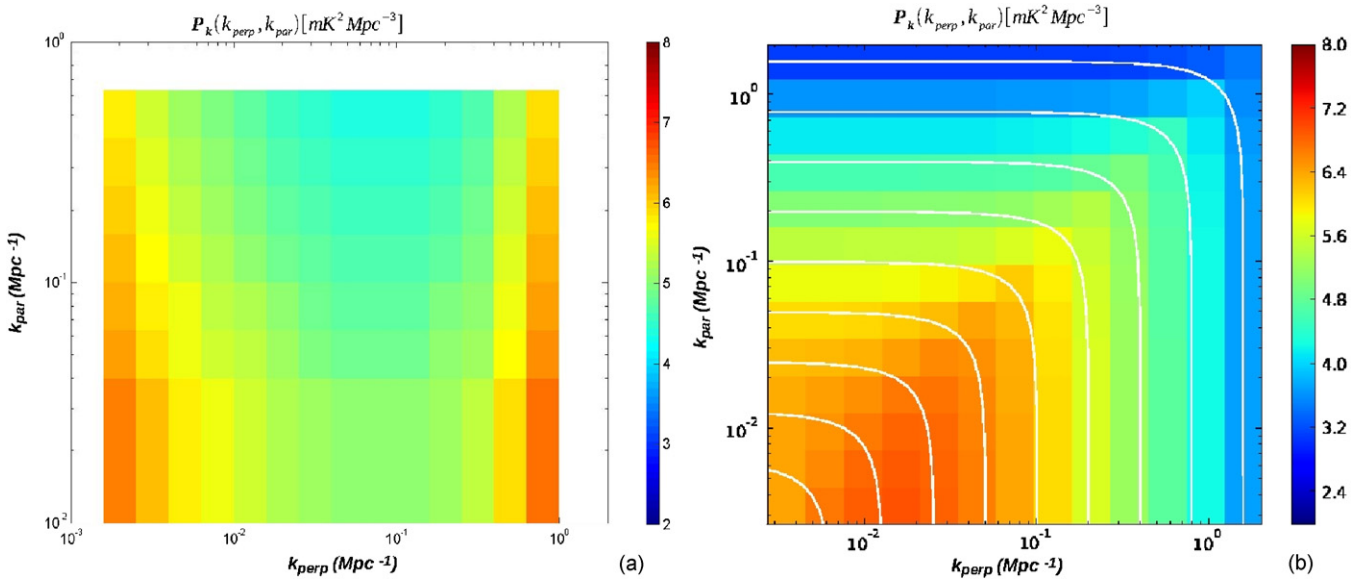

Figure 9. (a) Estimated 2D thermal uncertainty spectrum for 300 hr of integration with the MWA (Bowman et al. 2009). The thermal uncertainty is dependent on the shape and size of the binning operation in the 2D power spectrum. Here, the bins depicted in the figure are spaced at five per decade in both $k_\perp$ and $k_\parallel$. The MWA samples small $k_\perp$ much more densely than large $k_\perp$, but because the bins are logarithmic, the thermal uncertainty per bin reaches a minimum at intermediate values. (b) Theoretical 2D power spectrum of the H I 21 cm signal (Furlanetto et al. 2006) given by $P(k_\perp, k_\parallel) = (1 + 2\mu^2 + \mu^4)P(k)$, where $\mu = k_\parallel/|\mathbf{k}|$. Note that the quantity plotted here and in the following figures is $P(k_\perp, k_\parallel)$ in units of mK^2 Mpc^{-3}. The color scale is shown in $\log_{10} P(k_\perp, k_\parallel)$.

(A color version of this figure is available in the online journal.)

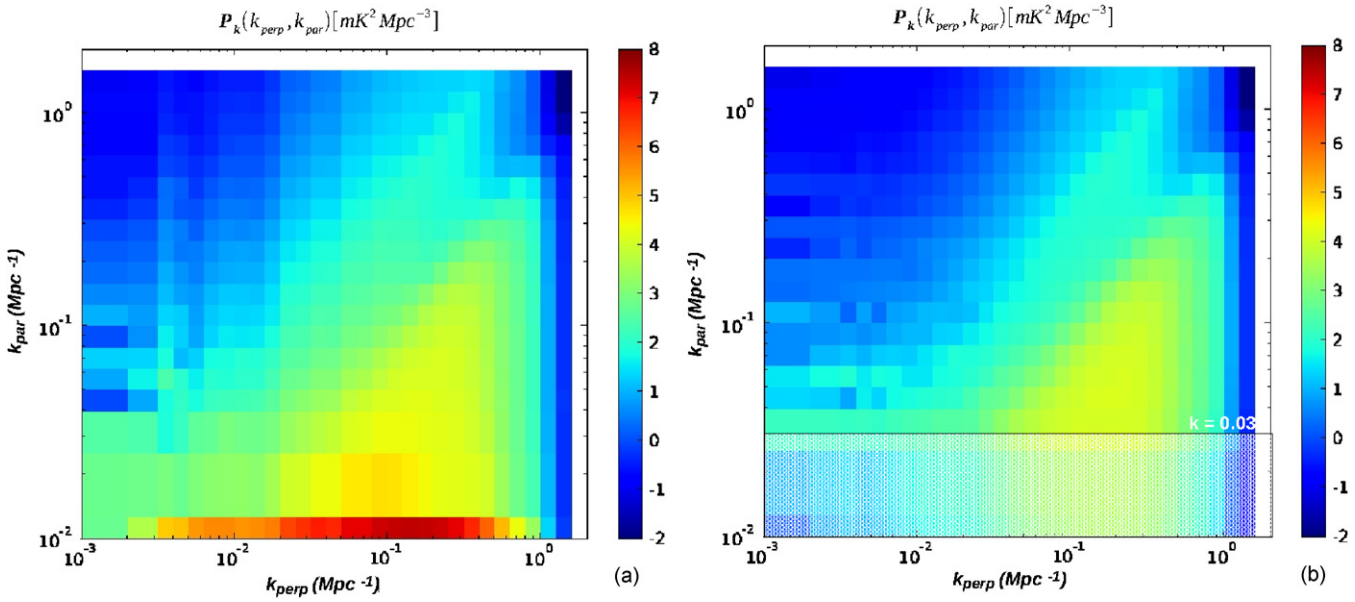

Figure 10. (a) 2D power spectrum of the GSM-subtracted residual image $I^{\rm res}(\vec{\theta}, \nu)$ made after subtraction of a foreground model with source position errors of $\sigma_\theta = 0.1$ arcsec. (b) Same as panel (a) but for the GSM+polynomial-subtracted residual image, $I^{\rm res}_{\rm polysub}(\vec{\theta}, \nu)$, that is produced after polynomial fitting and subtraction has been applied along each sight line. The shaded region in panel (b) corresponds to $k \leqslant 0.03\,{\rm Mpc}^{-1}$, where the polynomial fitting is expected to remove much of the structure. The color scale is shown in $\log_{10} P(k_\perp, k_\parallel)$.

(A color version of this figure is available in the online journal.)

the dominant structure is equivalent to the second harmonic in a *sinc* function. This results from the Fourier transform along the frequency of the relatively sharp boundary in the MWA UV coverage that occurs as any given UV radius transitions from receiving relatively sparse visibility sampling at lower frequencies to more dense sampling at higher frequencies as the core the MWA effectively expands in UV space (measured in units of wavelengths) from low to high frequencies.

In Figures 10 and 11, it is evident that there are modest differences between the residuals for the two cases of error that deviate from the commonality of the wedge-like feature. These are also easily accounted for by examining the subtraction model for each case. In the case of the source position errors, the residuals are the interference pattern from the beating of two foreground models that are nearly identical, with only the positions of each of the sources shifted by small angular offsets. In general, the interference pattern from this process at any given frequency will nearly cancel all power at large angular scales since the differences in phase of the perfect and the erroneous model visibilities are minimal at small UV (small $k_\perp$). The

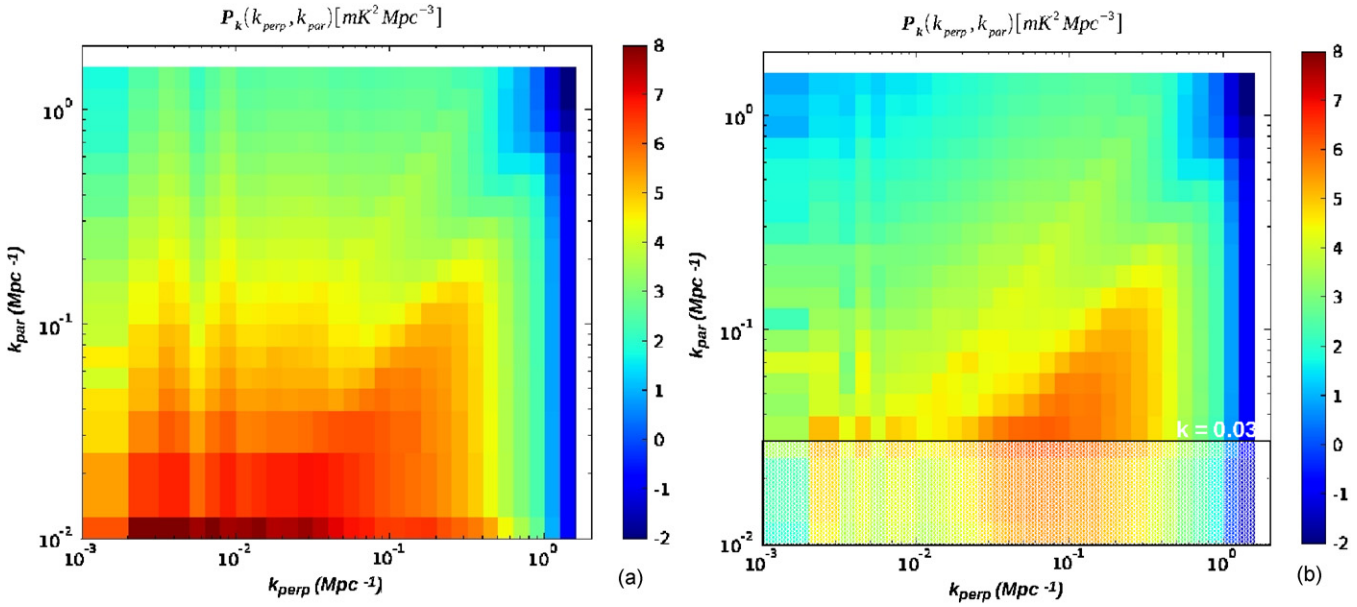

Figure 11. Same as Figure (10), but for the residuals due to calibration errors. In this case, the residual 2D power spectrum is shown for a fiducial calibration error level of $\sigma_a = 0.1\%$.

(A color version of this figure is available in the online journal.)

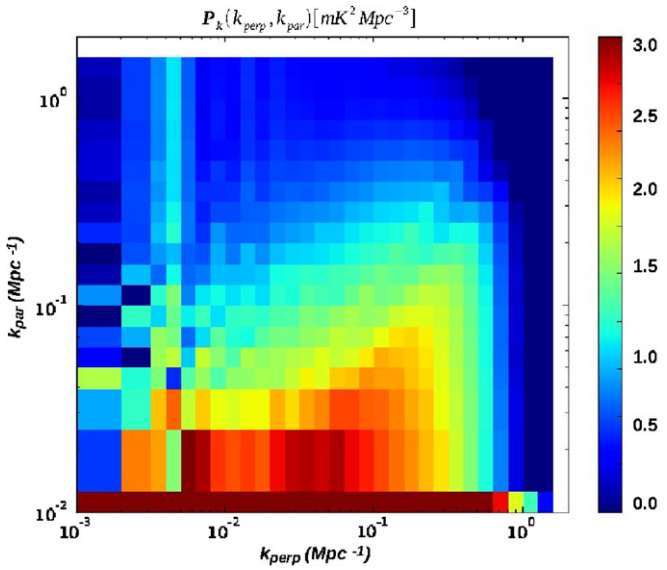

Figure 12. 2D power spectrum of the point-spread function (PSF) of the MWA-512 array for 6 hr observations. The color scale is in arbitrary logarithmic units.

(A color version of this figure is available in the online journal.)

phase errors will change little from one frequency to the next at small UV, whereas at large UV the phase differences are both much more pronounced and also change significantly between frequencies. Transforming this interference pattern into the 2D power spectrum results in the clean wedge shape in Figure 10.

On the other hand, the calibration errors assume that the perfect visibilities are multiplied by a Gaussian random gain field to produce the imperfect model that is subtracted from the perfect visibilities. Many independent baselines sample a given region in the UV plane at low UV, whereas relatively few baselines sample a similarly sized region at large UV. Hence, the magnitude of the calibration residual is also lower near the origin of the UV plane than at large radii. The calibration errors used in these simulations are assumed to be constant in frequency, but the frequency-dependent mapping of baseline length onto the UV plane causes variation in which baselines sample each UV cell as a function of frequency. This results in baselines with different calibration errors sampling the same UV region at different frequencies and, consequently, more line-of-sight structure is introduced into the residual calibration error maps at low-$k_\perp$ than for the case of source position residual errors. The prominent streaks extending from low- to high-$k_\parallel$ below $k_\perp \lesssim 0.05\,\mathrm{Mpc}^{-1}$ in Figure 11 are indicative of this distinction between the calibration error and source position error models.

3.4.2. Comparison to Expected Signal

Based on the fiducial theoretical H I signal shown in Figure 9(b), we see that the 21 cm signal is peaked in the lower-left corner of the 2D power spectrum at low-$k_\perp$ and low-$k_\parallel$, hence the wedge-like residual power from our source subtraction errors tends to bisect the expected signal structure. In Figure 9(a), however, it is evident that the most of the sensitivity of the MWA is concentrated around $\{k_\perp, k_\parallel\} \approx \{0.1, 0.05\}\,\mathrm{Mpc}^{-1}$, largely above the residual wedge. The secondary harmonic power evident in both Figures 10 and 11 does pass through the region of peak sensitivity for the MWA, but, fortunately, the power in the harmonic is an order of magnitude lower than in the primary wedge for both error models. Hence, our models produce residual 2D power spectra that together provide a consistent and clearly distinguished mask in order to exclude corrupted regions when averaging spherically binned power spectra or other statistics. It appears that searching for the H I signal at scales around $k_\parallel \approx 0.1\,\mathrm{Mpc}^{-1}$, which is fairly clean at lower $k_\perp$ values, would be highly advantageous for the MWA should bright source contamination be a problem in the final integrated maps. A "naively" derived spherically averaged power spectrum would contain significant contamination from the residual wedge at many scales that could be avoided by masking in the 2D $P(k_\perp, k_\parallel)$ domain. The cost of such an approach would be an associated reduction in sensitivity due to reducing the available input information for the averaging operation.

4. CONCLUSION

We have modeled two types of possible bright point source subtraction errors for the MWA, source position errors and baseline-dependent gain calibration errors. We find that the level of the source subtraction accuracy required for power spectral detection of the 21 cm signal is roughly comparable to the accuracy that would be required for direct imaging of the H I signal with the MWA (Datta et al. 2009). We have followed the historical development of redshifted 21 cm power spectra research in the literature and tested the source subtraction requirements in three types of power spectra: (1) angular power spectra, which we found to suffer large contamination due to the majority of the foreground power being confined to the angular plane of the sky, (2) a "naively" performed spherically averaged power spectrum that uses the entire observed data cube to achieve high S/N, but also suffered significant foreground contamination since regions of high contamination were not masked out, and finally (3) the 2D power spectrum, $P(k_\perp, k_\parallel)$, that yielded a common, clearly identifiable and localized wedge-like pattern of residual contamination at high-$k_\perp$ and low-$k_\parallel$, but otherwise left nearly half of the $k_\perp k_\parallel$-plane largely uncontaminated by residual foreground power.

The 2D power spectrum, in particular, addresses our second key question (as in Section 3.1), showing clear advantages for separating the residual contamination from the desired signal through distinct localization of the respective contributions in the $k_\perp$ and $k_\parallel$ plane. The results for both source position errors and residual calibration errors indicate that at $k_\perp \lesssim 0.05\,\mathrm{Mpc}^{-1}$, we are able to probe most of the $k_\parallel$ scales where the H I signal is dominant over the residual errors. In the 1D power spectrum we see dominant contribution from the residual errors around $\mathbf{k} \sim 0.1\,\mathrm{Mpc}^{-1}$, which can be probed in the 2D power spectra along the $k_\parallel$ axes.

Finally, our third key question (as in Section 3.1) was whether we might expect to build a template library of residual contamination errors in the power spectrum domain in order to facilitate interpretation of the final power spectrum. The results of this work indicate that it will indeed be possible. In both of the 2D residual power spectra for the source position error case and the gain calibration error case, the dominant wedge-like structure that contained the majority of the residual power was strongly coupled to the overall shape of the naturally weighted point-spread function for the MWA. Because the point-spread function is an easily calculated property of the array, this feature will be a valuable template during power spectrum estimation analysis. While the shape of such a feature will vary from one instrument to another, we anticipate that all upcoming 21 cm experiments should be able to use this result to improve the final stages of power spectrum analysis.

Previous studies have demonstrated that polynomial subtraction (or similar techniques) effectively remove confused point sources and diffuse foregrounds. Looking at the differences between our base GSM subtraction model and the GSM+polynomial subtraction, we find that the added step of polynomial subtraction is beneficial for the success of bright source foreground subtraction in the "naive" treatment of the spherically averaged power spectrum in the presence of source position errors. From our investigation of the 2D residual power spectra, however, it appears that the polynomial subtraction step for the source position errors only removes power in the already highly corrupted wedge-like structure, hence it likely would not provide much improvement over a masking operation. In the case of calibration errors, the polynomial subtraction step does remove much of the scattered power at high-$k_\parallel$ in the 2D residual power spectrum (as evident in Figure 11), so it would be helpful even after excluding the wedge region.

For the simulations included in this paper, we have performed the foreground subtraction of bright sources from a data set of 6 hr of observation in order to have the full effect of earth rotation synthesis, and then extrapolated to 300 and 5000 hr. However, the MWA may perform some of its bright source removal over much shorter timescales (~10 minutes or less) as part of its real-time calibration pipeline. The major implication for shorter timescale removal of the foregrounds would be to break our assumption when modeling source position errors that the position errors are constant for the entire observation.

We also made the assumption in this paper that each antenna's calibration errors are perfectly correlated for an entire 6 hr observation night, but uncorrelated between different observing nights. If the residual calibration error were instead perfectly random between every 8 s cycle of the real-time calibration processing planned for the MWA, then we estimate it could be possible to achieve the desired residual contamination noise level and detect the redshifted 21 cm H I signal from reionization with a significantly larger calibration error of $\sigma_a \approx 2.5\%$. We have not performed our detailed simulation under this assumption, however, nor have we used the exact parameters that will be employed for the real-time calibration pipeline of the MWA.

We would like to emphasis that similar bright source subtraction requirements can also be derived for other upcoming arrays, such as LOFAR and PAPER, as well as for future arrays like the Square Kilometer Array or a lunar array. But detailed simulations with the unique array specifications for each instrument would be required, which is beyond the scope of this paper. We expect to build on our present analysis in future work by exploring other arrays, addressing the modified scenarios described above, modeling sources more realistically with finite angular size and substructure (e.g., double-lobed radio sources), and including additional calibration issues such as wide-field gain calibration of the primary beam and ionosphere. Finally, we anticipating feeding real data and lessons from the MWA prototype system into our calculations to refine and improve the derived source and calibration model requirements.

A.D. and C.L.C. are grateful for support from the Max-Planck Society and the Alexander von Humboldt Foundation through the Max Planck Forshungspreise 2005. J.D.B. is supported by NASA through Hubble Fellowship grant HF-01205.01-A awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS 5-26555. The National Radio Astronomy Observatory is operated by Associated Universities, Inc., under cooperative agreement with the National Science Foundation. A.D. is grateful for support from NRAO through the graduate-intern program.

REFERENCES

Backer, D. C., et al. 2007, BAAS, 39, 967
Barkana, R., & Loeb, A. 2001, Phys. Rep., 349, 125
Blake, C., & Wall, J. 2002, MNRAS, 329, L37
Bowman, J. D., Morales, M. F., & Hewitt, J. N. 2006, ApJ, 638, 20
Bowman, J. D., Morales, M. F., & Hewitt, J. N. 2009, ApJ, 695, 183
Cornwell, T. J., Uson, J. M., & Haddad, N. 1992, A&A, 258, 583

Datta, A., Bhatnagar, S., & Carilli, C. L. 2009, ApJ, 703, 1851
Datta, K. K., Choudhury, T. R., & Bharadwaj, S. 2007, MNRAS, 378, 119
Di Matteo, T., Ciardi, B., & Miniati, F. 2004, MNRAS, 355, 1053
Di Matteo, T., Perna, R., Abel, T., & Rees, M. J. 2002, ApJ, 564, 576
Fan, X., Carilli, C. L., & Keating, B. 2006, ARA&A, 44, 415
Furlanetto, S. R., Oh, S. P., & Briggs, F. H. 2006, Phys. Rep., 433, 181
Gleser, L., Nusser, A., & Benson, A. J. 2008, MNRAS, 391, 383
Hales, S. E. G., Baldwin, J. E., & Warner, P. J. 1988, MNRAS, 234, 919
Harker, G., et al. 2009a, MNRAS, 397, 1138
Harker, G. J. A., et al. 2009b, MNRAS, 393, 1449
Harker, G., et al. 2010, MNRAS, 405, 2492
Hobson, M. P., & Maisinger, K. 2002, MNRAS, 334, 569
Jelić, V., et al. 2008, MNRAS, 389, 1319
Komatsu, E., et al. 2010, arXiv:1001.4538
Labropoulos, P., et al. 2009, arXiv:0901.3359
Lidz, A., Zahn, O., McQuinn, M., Zaldarriaga, M., & Hernquist, L. 2008, ApJ, 680, 962
Liu, A., Tegmark, M., & Zaldarriaga, M. 2009, MNRAS, 394, 1575
Lonsdale, C. J., et al. 2009, IEEE Proc., 97, 1497
McQuinn, M., Zahn, O., Zaldarriaga, M., Hernquist, L., & Furlanetto, S. R. 2006, ApJ, 653, 815
Mitchell, D. A., Greenhill, L. J., Wayth, R. B., Sault, R. J., Lonsdale, C. J., Cappallo, R. J., Morales, M. F., & Ord, S. M. 2008, IEEE J. Select. Top. Signal Process., 2, 707
Morales, M. F., Bowman, J. D., Cappallo, R., Hewitt, J. N., & Lonsdale, C. J. 2006a, New Astron. Rev., 50, 173
Morales, M. F., Bowman, J. D., & Hewitt, J. N. 2006b, ApJ, 648, 767
Morales, M. F., & Hewitt, J. 2004, ApJ, 615, 7
Morales, M. F., & Wyithe, J. S. B. 2009, arXiv:0910.3010
Myers, S. T., et al. 2003, ApJ, 591, 575
Paciga, G., et al. 2010, arXiv:1006.1351
Parsons, A. R., et al. 2010, AJ, 139, 1468
Peebles, P. J. E. 1993, Principles of Physical Cosmology (Princeton, NJ: Princeton Univ. Press)
Pen, U., Chang, T., Hirata, C. M., Peterson, J. B., Roy, J., Gupta, Y., Odegova, J., & Sigurdson, K. 2009, MNRAS, 399, 181
Shaver, P. A., Windhorst, R. A., Madau, P., & de Bruyn, A. G. 1999, A&A, 345, 380
Thomas, R. M., et al. 2009, MNRAS, 393, 32
White, M., Carlstrom, J. E., Dragovan, M., & Holzapfel, W. L. 1999, ApJ, 514, 12
Zaldarriaga, M., Furlanetto, S. R., & Hernquist, L. 2004, ApJ, 608, 622